**CS 59000/STAT 46700 – Topics in Data Science**

**Homework 1**

1. Given two points **a**=(6,5,4) and **b**=(3,2,1) calculate the following measures between **a** and **b**
2. Euclidean Distance
3. Manhattan Distance
4. Minkowski distance (choose order of the norm p=3)
5. Consider a dataset 2, 4, 6, 8, 8, 10, 12, 14. Express the data in standardized form using
6. Minmax approach in (0,1)
7. Z-score
8. Calculate the entropy of a weighted six-sided dice such that three sides of the dice have 1/6 chances of facing, two sides of the dice have a 1/12 chance of facing up, and one side has a 1/3 chance of facing up.
9. For the dataset given below, find the approximate entropy H (Passed). This data describes whether students pass or not (Y for yes or N for no), based on their past CGPA scores (H for high, A for average, and L for Low) and whether they prepared or not (Y or N).

|  |  |  |
| --- | --- | --- |
| CGPA | Prepared | Passed |
| H | N | Y |
| H | Y | Y |
| A | N | N |
| A | Y | Y |
| L | N | N |
| L | Y | Y |